1948

A critical analysis of scores, accumulated from the presentation of the Wechsler-Bellevue Intelligence Scale to adolescent and adult deafened individuals

Loraine Doris Meier

WASHINGTON UNIVERSITY

Central Institute for the Deaf

A CRITICAL ANALYSIS OF SCORES, ACCUMULATED FROM THE PRESENTATION OF THE WECHSLER-BELLEVUE INTELLIGENCE SCALE TO ADOLESCENT AND ADULT DEAFENED INDIVIDUALS

by

Loraine Doris Meier

A dissertation presented to the Board of Graduate Studies of Washington University in partial fulfilment of the requirements for the degree of Master of Science in Education

June, 1948

Saint Louis, Missouri

TABLE OF CONTENTS

TABLES

FIGURES

Chapter I

INTRODUCTION

Language tests of intelligence have not been used satisfactorily with the deaf. When they have been used, they have indicated mental retardation. In order to get a measurement of intelligence, it has been necessary to substitute performance tests. Many psychologists have questioned the validity of performance tests as measures of intelligence especially at adolescent and adult age levels. This study of the Wechsler-Bellevue Intelligence Scale is one that incorporates both the verbal and performance sections with a statistical relationship computed between them. The question of whether a portion of such a test alone could be used depends largely upon the concept of the nature of intelligence.

As defined by Wechsler:

> Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment.[1]

It is an aggregation or sum total of abilities, such as a retentive memory, mechanical dexterity, and ability to reason, which can be differentiated qualitatively, although not completely independent of one another. It is global because it is representative of the individual's behavior as a whole. Intelligence is evaluated by measuring these abilities, but a final definition is not the equivalent of their sum.

In the construction of the Bellevue Scale, Wechsler

---

[1]Wechsler, David, The Measurement of Adult Intelligence, Baltimore: The Williams and Wilkins Co., 1944, p. 3.

assumed this hypothesis:

> Any practical definition of intelligence must be fundamentally a biological one in the widest sense of the term.[2]

The quantity measured by intelligence tests is more than the single factor "g", defined simply as mental energy. Rather, it is the ability to apply this energy in situations that have "content and purpose as well as form and meaning". The scale used in this study does not measure all that goes to make up general intelligence, for there are at the present time no tests available which are capable of doing this.

It has been frequently suggested and test results support the claim that some individuals are more fairly tested by non-verbal tests of intelligence than by verbal tests; indeed, both types should be administered to obtain a meaningful picture of an individual's mental capacity. Brody, in studying comparable tests of verbal and non-verbal reasoning, summarized their application to developmental problems with this statement:

> Such tests will serve to depict more completely an individual's mental ability, and to examine more fairly in those cases where the language factor has to be minimized or controlled.[3]

In any attempt to test the deaf whether they be children or adults, a primary difficulty centers about the question of how the large discrepancy between the results of

---

[2] Ibid., p. 11.

[3] Brody, Leon, "Comparable Tests of Verbal and Non-Verbal Reasoning: Their Construction and Application to Developmental Problems", The Journal of Educational Psychology, 180:194, 1940.

performance tests and tests involving language should be interpreted. This variance is significant since it means that non-language and performance tests do not necessarily measure the kind of ability required for success in school work dependent upon language.

Clinicians have for many years used performance tests in place of language tests, such as Pintner and Paterson's Scale of Performance Tests, Grace Arthur's Point Scale of Performance, the Goodenough Test, the Cornell-Coxe Performance Ability Scale, and the Advanced Performance Series. They have been applied when verbal difficulties are present, but their test material often involves a speed-of-motor performance which may, at the upper age levels, be confused with intellectual ability.

In 1943, Heston's factor analysis of representative performance tests bore out a result previously mentioned, that no single general ability can be found common to all the tests.[4] That is simply another way of stating that intelligence tests cannot measure all of intelligence. In fact, at different ages they measure different portions of it. Also, as the individual grows older, tests become less and less effective measures of his global intelligence. Failure to accept this premise has led to a great deal of confusion in interpreting test results. When intelligence levels are given as mental ages, then the evaluation of adult intelli-

---

[4]Heston, Joseph C., "A Factor Analysis of Some Clinical Performance Tests", The Journal of Applied Psychology, 135:149, 1943.

gence becomes most acute. To quote Wechsler:

> The basic reason a mental age of twelve at twelve does not mean the same thing as a mental age of twelve at thirty, is that the measured abilities and hence the M.A. scores represent different portions of the subjects' respective total intelligence. At age twelve the tests are capable of tapping far more of the individual's capacities than at thirty.[5]

From nursery school to adolescence, the mental capacities of the deaf child are compared with the hearing. The time soon comes when the boy completes grade school or, perhaps, graduates from high school. He is ready to take his place as a useful citizen of his community. To further his opportunities and to better his chances for success, he seeks vocational guidance either at school or from a social service agency. Psychometrists put him through a series of intelligence and aptitude tests, the former of which have been standardized for the most part on children. Yet, he is an adult. Can a test with norms, obtained from measurement of children, serve as a means for classification in industry where information about intelligence is useful? Wechsler believes not. In 1939, he made a careful analysis of existing tests and incorporated a final choice of their subtests into four related scales. His point scale supplies the demand for an adult test for the hearing, but with what is the intelligence of the deaf to be measured?

From 1889 to 1915, practical psychological obser-

---

[5]Wechsler, David, The Measurement of Adult Intelligence, Baltimore: The Williams and Wilkins Co., 1944, p. 12.

vations of deaf children were made, in the attempt to use the data as criteria for admission to schools and comparison with hearing children.

From 1915 to 1928, it was the general feeling of those who carried on psychological measurement that the deaf child, by reason of the handicapping effects of his loss of hearing, was not equivalent mentally to the average child his age. Supposedly, he was retarded, in almost every measure of ability when compared with the hearing. This view was accepted without question by the majority of Pintner's followers. Still, it was not universal. The debatable question of normalcy with regard to a deaf child's intelligence encouraged the development and use of other tests, besides those applied by Pintner.

Results from 1930 to the present date show a trend toward a normal distribution of intelligence for the deaf child, using various types of performance tests as measures. The discrepancy in recent results as compared with Pintner's can be attributed to: (1) the use of tests that were non-verbal in performance but which required some language in administration, such as "Do this as quickly as you can"; (2) the quality of pupils in the schools used for testing. Before psychological measurements were developed, many children who did not develop speech because of low mentality were placed in state institutions for the deaf and included in the results. (3) Pintner's results were based largely on group tests while recent results are derived from individual tests.

In the examination of all of the available tests

used to measure the intelligence of the deaf, it was learned that the Pintner-Paterson Performance Scale, designed to measure the mental abilities of young deaf children, could be presented to adolescents, as could also the Advanced Performance Series, a scale of tests standardized at Central Institute for the Deaf. The norms for both measures are available up to age sixteen, but is a person justified when computing the intelligence of an individual, forty-five years old, to use the same statistical tools that one would employ for a sixteen year old?

Finding no adequate measure of intelligence usable for deaf adults, the writer selected the Bellevue Scale, reputed to be an effective measure of the intelligence of hearing adults, to be used with the deaf. How would their scores compare with those of the hearing on whom the test had been standardized? Could the scale be presented to the deaf in its entirety, to include both verbal and performance items? To what extent would education influence the intelligence quotients of the deaf? How do the Bellevue intelligence quotients compare with those of other mental tests given to deaf adults during the course of their childhood education? How are the deaf adjusting vocationally to changing circumstances?

The present investigation has been made to obtain information concerning the application of the Bellevue Scale to the deaf; to compare scores of the deaf with Wechsler's norms for the hearing; to analyze the verbal subtests with respect to the difficulty of the test material; and to determine the influence of education on the performance quotient, verbal quotient, and total scale.

Chapter II

THE NEED FOR AN ADULT INTELLIGENCE TEST

A need for adult intelligence tests grew out of World War I when the United States Army sought a means of classifying and selecting its officers from the soldier draft. About the same time, Terman made improvements in the Stanford Revision of the Binet Scale by including tests at the adult level. He recommended:

> using a basic chronological age of sixteen years, if one is testing an adult. The assumption here is that this represents normal adult intelligence and that any stage of mental development, either above or below this, should be judged by this criterion.[1]

The revised Binet was shelved so far as the Army was concerned. Its usefulness as an individual test yielded to the urgent demand for group tests where a large number of men could be examined in a relatively short time. This want for an adequate measure of adult intelligence was fulfilled by psychologists devising the Army Alpha and Beta group tests. They were given to more than a million men and constituted the most thoroughly standardized group tests yet devised.[2] In the opinion of Wechsler, the Alpha test still remains the best instrument of this type. Its merit rests on the dictum that the examiner can make a rapid classification of a large number of subjects. However, it does not offer a means for an individual

---

[1] Boynton, Paul L., Intelligence: Its Manifestations and Measurement, New York: D. Appleton and Co., 1933, p. 48.

[2] Wechsler, David, The Measurement of Adult Intelligence, Baltimore: The Williams and Wilkins Co., 1944, p. 14.

diagnosis.

A number of adult group tests were devised following World War I. In the main, they were patterned after the Army Alpha. With the available individual intelligence examinations, the Army found through experience that they were generally inadequate for use with adults. Reports from a study with the Stanford-Binet, on some four hundred unselected subjects, led to the discovery of a mental age adult norm three years lower than that given by Terman. Thus, the average mental age for the American soldier turned out to be 13.08 years, instead of sixteen years as originally claimed by Terman. A widespread controversy resulted from this supposed "decrease" in the intellectual status of the adult male population. Test results showed an estimate of forty percent below a mental age of twelve. If Terman's distribution of I.Q.'s is again considered and an adult I.Q. computed using a basic mental age of sixteen, the result would be that forty percent would have an I.Q. of seventy-five, whereas only two or three percent of children are classified so low.

In 1937, Terman and Merrill published another revision of the Stanford-Binet in which they lowered the average mental age to fifteen years. But, as with their first revision, the tests were again standardized without benefit of adult responses. Individuals over eighteen years of age were not included in the experimental group. Hence, the "true" mental age of the average adult still remained to be found.

The Binet and other children's scales continued to be used for the testing of adults until 1939. Despite abundant

criticism, this practice was carried on for several reasons: (1) there was no satisfactory individual adult examination available; (2) the children's tests furnished the desired information, requested at the moment; (3) the continued employment of these tests seemingly made their users more oblivious of their shortcomings; (4) the standardization of a new adult scale would have presented great difficulties.

Many of today's tests for adults do not possess the essential characteristics needed in measuring intelligence. Firstly, they are not standardized on a sufficient number of cases and most of them were never standardized on adults. Wechsler backs up his statement with the following illustration to prove how serious the situation is in this country. These are the individual tests most commonly used:

Table I

Individual Tests Most Commonly Used for the Measurement of Adult Intelligence

Test	Number of Adults
Stanford-Revision of the Binet	62
Kuhlmann-Binet	Not specified
Yerkes Point Scale	73
Pintner-Paterson Performance	0
Army Performance Scale	260
Terman-Merrill Revision of the Stanford-Binet	0

The authors of these tests computed norms for them but no published studies have confirmed the data. Terman speaks of the Stanford Revision of the Binet having been given to one hundred fifty migratory workers but the subjects' scores do not seem

to have been used in the final standardization.[3]

Secondly, Wechsler claims that much of the test material is unsuitable.

> To ask the average adult to say as many words as he can think of in three minutes, or to make a sentence of the words "to asked paper my teacher correct I my" and assume that he will be either interested or impressed, is expecting too much.[4]

The normal person will start wondering what good the test will do him, as an individual.

Many of the tests, originally standardized on children and used with adults, lay too much stress on speed rather than accuracy. This emphasis varies widely in different scales, and the relation of speed to intelligence is still a question upon which many psychologists do not agree.

> Here, again, conclusions differ diametrically: on one side are those who find no connection between quickness and intelligence, such as Bernstein and Highsmith. On the other side are those who assert that speed may often be taken as an indication of intelligence. Ruch and Koerth found a high correlation between speed and power in the Army Alpha test. Peak and Boring make the statement "Speed of reaction is an important, and probably the most important factor in individual differences in the intelligent act". McFarland has shown that an individual's speed is a constant factor throughout a wide range of mental performance, while Kelley suggests that "mental speed" may be an "independent trait". Finally, Thorndike concludes that the correlation between speed and level of intelligence is so low that not much weight should be attached to the former except where it measures speed of learning.[5]

---

[3]Terman, L., The Measurement of Intelligence, Boston, Mass.: Houghton Mifflin Co., 1916, p. 54-55.

[4]Wechsler, David, The Measurement of Adult Intelligence, Baltimore: The Williams and Wilkins Co., 1944, p. 17.

[5]Schieffelin, Barbara, and Schwesinger, Gladys C., Mental Tests and Heredity, New York: The Galton Publishing Co., Inc., 1930, p. 38.

Wechsler admits there is a high correlation between speed and accuracy and that in the case of children, speed is often the better measure for it correlates higher with other global ratings of intelligence. With adults, however, speed alone is not the better measure. One needs to consider:

> the possible influence of the different attitude which adults take toward set tasks or test situations.[6]

Time scores can be used with adults provided one evaluates and interprets them in a different light than that used with children.

According to Wechsler, the most fundamental reason of all why tests standardized on children cannot be used on adults is that adult intelligence cannot be evaluated in the same terms as those generally employed in defining juvenile intelligence.

The Wechsler-Bellevue Intelligence Scale has been designed to meet the shortcomings ascribed to intelligence tests available for adult measurement. In the scale, the age factor was taken into account by establishing norms for all ages up to sixty. The actual number of cases tested was 3499, and from this sampling, 670 cases were used in the age distribution of subjects between seven and sixteen years, and 1,081 cases for ages seventeen to sixty-nine years. The subjects were chosen mostly from the city and state of New York, and were matched against the total population of the United

---

[6] Wechsler, David, The Measurement of Adult Intelligence, Baltimore: The Williams and Wilkins Co., 1944, p. 18.

States. The authors tried to obtain individuals from all walks of life, and after much consideration, felt that:

> the most unbiased method of selection would be a sampling based upon the occupational distribution of the country's adult population.[7]

The United States Census for 1930 furnished the necessary data.

Wechsler's aim in constructing the Bellevue Scale was not to produce a set of brand new tests, but by drawing heavily on the experience and sources of others, to select such a combination of them as would best meet the requirements of an effective adult scale. In presenting the Bellevue Scale to fifty deaf adults, the examiner found that it definitely held the interest of the subjects. This was especially true in the case of the Information, Comprehension, Arithmetical Reasoning, Digit Span, Picture Arrangement, and Block Design subtests. The questions or problems in the first three tests mentioned above touch upon commonplace situations or involve practical calculations. The subjects looked upon them as a task worthy of an adult's consideration. They found no reading difficulties whatever as the items were clearly worded. Their responses to the Digit Span test were surprising, for they would call back the numbers as though playing a game. To them, the situation presented a challenge different from any they had experienced in previous tests. They readily entered into the "spirit" of the picture series because their attention centered on the clever caricatures of the items. The Block Design test, a common feature of almost every perform-

---

[7] Ibid., p. 108.

ance series, has appealed to the majority of them since childhood.

In the Bellevue Scale, credit for correctness of response does not depend upon the individual's ability to manipulate words, a characteristic of material much employed in children's tests. Rather, what makes the material so very suitable for adults is its leniency in scoring. Half of the Verbal subtests - Comprehension, Similarities, and Vocabulary - allow the individual to score when he only partially comprehends the meaning of a word or sentence. The other three subtests, however, demand a precise accuracy of response. That is expected when one considers the nature of the tests, such as, possession of knowledge needed to reason out an arithmetic problem.

Older subjects do not consider speed important. They would rather perform a task well, such as placing the pieces of a puzzle together correctly, and in so doing, take their time, rather than hurry and do careless work. They will think out a problem to get a clear mental picture of what they are supposed to do before they make any attempt to solve it. The subtests - Picture Arrangement, Object Assembly, Block Design, and Digit Symbol - depend as much upon accuracy as they do upon speed. When using a test for diagnostic purposes, an examiner needs to interpret the way in which a subject responds to a problem situation. How fast he works becomes a secondary factor to his approach and performance of the task in general.

Wechsler does not deny the value of the mental age concept altogether but only points to its unavoidable limita-

tions.

> The most important of these limitations is that the M.A. method of defining intelligence cannot logically be used to define levels of intelligence higher than that obtained by that age group beyond which M.A. scores cease to increase with chronological age.[8]

The usual formula for calculating an individual's intelligence quotient is by dividing a subject's mental age by his chronological age. Bearing in mind the fact that the M.A. and the C.A. are alike in that they are both converted test scores measured in identical units, Wechsler defines an intelligence quotient in terms of what it actually represents to him.

> An intelligence quotient is the ratio between a particular score which an individual gets (on a given intelligence test) and the score which an average individual of his life age may be assumed to attain on the same test, when both scores are expressed in the same notation (e.g., in terms of months and years).[9]

In a shortened form, the I.Q. may be stated as follows:

$$\text{I.Q.} = \frac{\text{attained or actual score}}{\text{expected mean score for age}}$$

To obtain an individual's I.Q., psychologists have generally adapted the method of using the "highest C.A. beyond which the observed M.A. scores cease to increase" as the divisor. Different authors have set this age anywhere from fourteen to eighteen years, depending upon the particular scale employed. But, the fixing of this assumed age at a set point has served to destroy the meaning of the I.Q. altogether. The assumption is that M.A. scores remain constant throughout adult

---

[8] Ibid., p. 22

[9] Ibid., p. 24

life. From data obtained with the Bellevue Intelligence Examinations for ages seven to sixty-five, Wechsler has shown, in Figure I, that actually the curve from age sixteen onward is not a straight line parallel to the C.A. axis as the above assumption declares, but rather a decreasing one, falling off regularly with increasing age.

Beginning at an age varying from fifteen to twenty-two, all scores of mental ability begin to fall off. The successive increments by which the test scores have increased with advancing chronological age have diminished, and after age twenty-five, they have vanished altogether. After age thirty-five, the decline becomes so appreciable that the use of a single denominator for calculating I.Q.'s "will introduce serious errors". Wechsler further adds:

> To calculate an I.Q. for a man of sixty by dividing his M.A. score by fifteen, is as incorrect as to obtain an I.Q. for a boy of twelve by dividing his M.A. score by fifteen. Nevertheless that is precisely what the general practice has hitherto been. All adult I.Q.'s have been almost universally calculated by using some assumed constant denominator. The effect of using a single denominator for calculating I.Q.'s of adults is to destroy the very basic concept of the I.Q.[10]

---

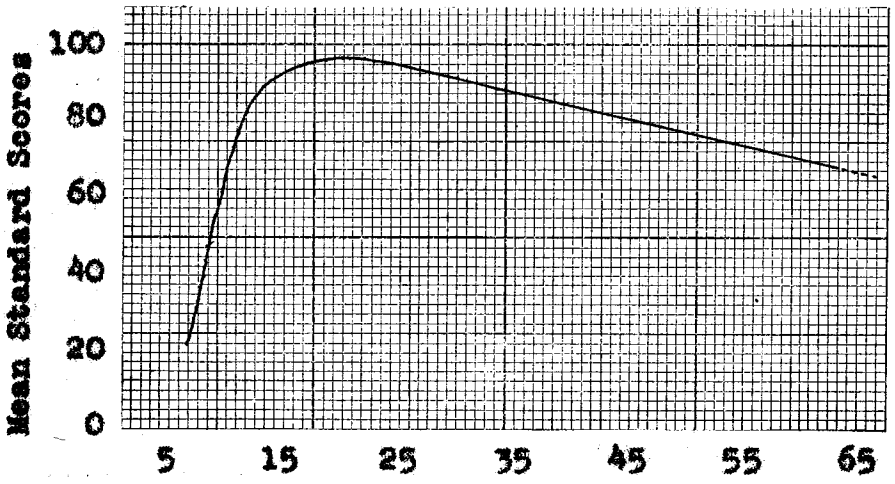
[10] Ibid., p. 30.

Figure I

CURVE OF MENTAL GROWTH AND DECLINE

Bellevue Full Scale[11]

Ages 7-68



[11] Ibid., p. 29.

Chapter III

DESCRIPTION OF THE TESTS

The Bellevue Intelligence Examination consists of eleven subtests combined to form four separate intelligence scales, listed as follows:

1. An INDIVIDUAL ADULT EXAMINATION for ages sixteen to sixty. This scale includes ten tests but permits a minimum of seven tests, depending upon their suitability to the subject. If ten tests are given plus the vocabulary test, as an alternate, the scale is known as the "Full Scale".

2. An ADOLESCENT Scale for ages ten to sixteen. It consists of the same tests but separately standardized. It too is a Full Scale if ten tests are presented, in addition to the vocabulary as an alternate.

3. A PERFORMANCE Scale consisting of five tests (test 6 to 10 inclusive).

4. A VERBAL Scale consisting of five or six tests (test 1 to 5 and the vocabulary as an alternate).

Under usual testing conditions, the I.Q. is calculated from the subject's total weighted score on the Verbal and Performance tests only. Vocabulary does not enter the I.Q. computations because it holds a special position in the Bellevue Scale. It serves as an alternate for any verbal subtest if one of them cannot be given. Also, its stability provides a basis of comparison for the impairments of the other tests.[1]

Rapaport, in his statistical evaluation of the Bellevue Scale, found that the two major test divisions - Verbal

---

[1]Rapaport, David, Diagnostic Psychological Testing, Chicago: The Year Book Publishers, Inc., 1945, p. 87.

and Performance - could be broken down into further subgroupings. He emphasizes that the four subtests - Vocabulary, Information, Comprehension, and Similarities - are essentially verbal and should be considered the Verbal part proper; Digit Span and Arithmetic diverge in nature to a great extent from the other verbal tests as they measure the related functions of attention and concentration; in the Performance part, picture arrangement and Picture Completion are dependent to a large degree on visual organization, but motor activity is not essential to their achievement; Block Design, Object Assembly, and Digit Symbol are chiefly visual motor coordination tests.

Only one of the Verbal subtests, Arithmetic, is a timed one. On all its items there is a time limit and on two items, additional credit is given. On the Performance part, each subtest has time limits, and on all the subtests except Picture Completion the subject may obtain added time credits. The subject can obtain full and half-credits on the Comprehension, Similarities, and Vocabulary items; Picture Arrangement, Object Assembly, and Digit Symbol make allowance for partial scores, while Picture Completion and Block Design give credit only for an accurate performance.

In the following descriptions of the individual subtests, it will be noted that the testing situation provides a contrast between the Verbal and Performance parts and between timed and untimed subtests.

Test 1. Information

Wechsler made use of information items when devising his intelligence scale both from a practical consideration of

their ease of scoring, and from the support which the Army Alpha Examination gave. Its data proved, by high correlation with various estimates of intelligence, that an information test can be a very good indication of a person's intellectual capacity. The value of such a test depends upon the "good" characteristics of the questions themselves: how well they can discriminate at the various levels of intelligence, the order of difficulty in their presentation, and the fact that the items should call for the sort of knowledge that an average individual with average opportunity may be able to acquire for himself.[2]

The Information subtest consists of twenty-five items. Their subject content includes general questions, such as "How many pints make a quart? When is Washington's birthday?", and more specific ones, as "What does the heart do? What is an Habeas Corpus?". All of the items, except the Habeas Corpus question, can be answered by a simply stated fact.

The test is scored on the basis of a plus or minus for each response. The maximum number of correct responses is the total of twenty-five.

Test 2. General Comprehension

This is a test of common sense. Its success depends upon the possession of a certain amount of practical information and a general ability to evaluate past experience.[3]

---

[2]Wechsler, David, The Measurement of Adult Intelligence, Baltimore: The Williams and Wilkins Co., 1944, p. 78.

[3]Ibid., p. 81.

The subtest consists of ten items and two alternate questions. They were selected from an original thirty questions, some of which were chosen from the Army Alpha, or from those tests discussed in the Army Memoirs.

The examiner presents the first question, "What is the thing to do if you find an envelope in the street, that is sealed and addressed and has a new stamp?", and proceeds to record the subject's response verbatim. After presenting the complete list of questions, he matches the subject's responses against sample answers given on the scoring sheets. The responses are scored either two, one, or zero, depending upon the extent of generalization and the quality. A complete response to an item is credited with a score of two, an incomplete response with a score of one, and a failed response with a score of zero. The total score is the sum of the credits on the ten questions.

Test 3. Arithmetical Reasoning

> Most intelligence scales include items calling for arithmetical reasoning in some form. The inclusion of such items is fully justified: Arithmetical Reasoning tests correlate highly with global measures of intelligence.[4]

The Arithmetic subtest of the Bellevue Scale consists of ten items. With the first eight "daily life" problems, the examiner gives these instructions: "Read this one to yourself and do the problem mentally".

> Arithmetic is considered here a test of concentration, because to pass the items of this subtest...the

---

[4] Ibid., p. 82.

> average person of our civilization must utilize patterns of arithmetical relations ingrained in him.[5]

The items require the four basic calculations for their solution i.e.

No. 1 - a simple addition (4-5)
No. 2 - a simple subtraction (10-6)
No. 3 - a subtraction over a ten limit (25-8)
No. 4 and 5 - each a simple division (36÷4; 24÷3)
No. 6 - a simple multiplication and a subtraction over a ten limit (50÷7x2)

The subject solves the problems until he fails three successive examples. The items of the Arithmetic subtest are timed, but only the last two merit an additional credit. Three problems are to be answered in fifteen seconds, three in thirty seconds, two in one minute, and two in two minutes.

The remaining two items are presented with these instructions: "Now read this one aloud and do the problem mentally". Timing begins after the subject has finished reading the problem. If the answer is given within forty seconds, the subject receives an additional credit of one; if given within fifteen seconds, two more points are added to his score.

The test is scored by allowing one point for each item answered correctly within the time limit, and by giving additional credit of one for time on problems nine and ten.

Test 4. Memory Span for Digits

This test is a very familiar one, being used in almost every scale of intelligence. The original Binet Scale

---

[5]Rapaport, David, Diagnostic Psychological Testing, Chicago: The Year Book Publishers, Inc., 1945, p. 195.

included it. Although it has as many bad features, being a poor test of general intelligence, as good features, it forms a part of the Wechsler-Bellevue Intelligence Scale because the repetition of digits backward and forward is often of diagnostic significance.

Digits forward and backward are given as two separate parts. The first requires repetition of a series ranging from three to nine digits in length; the second requires repetition backward of a series from three to eight digits in length.

Beginning with three digits, the examiner directs the subject to watch carefully, and when he finishes calling out some numbers, to repeat them right after him. If the first series is repeated correctly, the examiner marks it plus and presents the second group of digits - four numbers. If the subject fails the series, he is given a second trial on a series of equal length. Failure on both trials discontinues the presentation of future series. The highest number of digits repeated without error on either of the two trials is the test's score.

Digits backward is administered in like manner, except that the subject reverses the order of presentation. The score is the highest number of digits correctly reversed.

The total score on the two parts of the Digit Span subtest is the sum of the number of digits repeated forward and backward.

The Digit Span subtest was presented to the deaf with much hesitation, the examiner being convinced that another fac-

tor besides rote memory would influence the ability to repeat back the numbers to the examiner. An individual's lipreading efficiency requires concentrated effort and the utmost attention. This factor, alone, would play a large part in the test's success. Hence, Memory Span for Digits is considered here a test of concentration. Considering the fact that the lipreading of numbers is, on the whole, an easier accomplishment than grasping the understanding of isolated sentences, the test was included rather than omitted from the scale battery. Too, a desire to find out whether or not such a test could be given to individuals so handicapped by a loss of auditory sensation spurred the examiner on.

Test 5. Similarities

As claimed by Wechsler, this test turns out to be among the best of the entire battery.[6] It was omitted from his original selection of tests, as he felt it would be influenced by a language factor. Research showed that it was possible to increase the difficulty of the items without employing unfamiliar words. The Similarities Test is easy to give, has a definite interest appeal, and is especially important for the light it throws upon the logical character of the subject's thinking processes.[7]

The Similarities subtest of the Bellevue Scale consists of twelve items. The examiner presents them to the sub-

---

[6] Wechsler, David, The Measurement of Adult Intelligence, Baltimore: The Williams and Wilkins Co., 1944, p. 85.

[7] Ibid., p. 86.

ject with these directions: "I am going to name two things which are the same or alike in certain ways and I want you to tell me in what way they are alike. For example: In what way are an orange and a banana the same?"

The subject's reply is recorded verbatim, for the method of scoring distinguishes between superior and inferior responses by allowing different credits for each. The list is continued until four successive pairs are completely failed. The examiner matches the responses against sample answers and scores accordingly - either two, one, or zero, depending upon the degree and quality of the generalizations.

Test 5A. Vocabulary

A vocabulary test was included in the scale for several reasons: (1) a vocabulary test is reputed to be an excellent measure of a person's general intelligence. It is a popular belief that the number of words a man knows is a measure of his learning ability, his fund of verbal information, and of the general range of his ideas.[8] (2) A vocabulary test has qualitative possibilities. The subject will not only give his definition of a word, but, in so doing, will enable an evaluation of his thought processes and relate some information concerning his cultural background which will have partly influenced that response. To be sure, a vocabulary test is dependent upon educational opportunities.

> The intellectual blocking that has afflicted mental testing for a quarter century is susceptibility to

---

[8] Ibid., p. 99.

the influence of schooling.[9]

Also, it is seemingly unfair to persons with a language handicap. The authors found this last apprehension to be unfounded, and recommended the inclusion of the vocabulary test as a "regular" in their battery.

The subtest consists of forty-two items, taken from one of the Funk and Wagnall's Standard (School) Dictionaries. Arranged in order of difficulty, beginning with "apple" and ending with "traduce", the words seem to discriminate well between different levels of intelligence. The examiner records the responses verbatim. If the subject shows that he knows what a word means, he is given full credit, and no penalty is placed upon inelegance of language. If the examiner is in doubt concerning the correctness of the word's meaning, he matches it against the acceptable definitions included in the manual and scores accordingly. Items satisfactorily passed are given a raw score of one; in cases where the subject seems to know only vaguely what a word means, his response merits a raw score of one-half. Final score is the sum of partial and full credits.

Test 6. Picture Arrangement

This performance subtest consists of a series of pictures similar to the short comic strips found in the daily papers. The examiner presents the material to the subject in a disarranged fashion with the directions to put the cardboard squares in the right order so that the pictures on them make a meaning-

---

[9]Stoddard, George D., The Meaning of Intelligence, New York: The Macmillan Co., 1943, p. 112.

ful story sequence.

First used by a Frenchman, named DeCroly, in 1914, a test of this sort was later incorporated into the Army Performance Scale. Yet, it didn't find too much favor in this country because of difficulties in scoring, as well as in getting up good sequences. In 1934, picture series came into vogue again and were included in Cornell and Coxe's Scale.

Wechsler chose a sample demonstration and six items for his test. Three series were adapted from the Army Group Examinations and four new ones were selected from Soglow's "King" series. For example, item (5) includes:

> six pictures about a fishing experience, showing the "Little King" waiting for a catch, getting one, waiting again, getting another catch, calling down toward the water, and his chamberlain emerging in a diver's suit from the water where he has been putting fish on the King's line.[10]

Preference was given to the picture series that met certain criteria: interest of content, probable appeal to subject, ease of scoring, and discriminating value. The author's aim was to choose situations from the American scene. Success on this test depends upon how well the subject gets the "drift" of the story before he is able to effectively rearrange the pictures.

The examiner records the subject's actual arrangement by giving attention to the code letters on back of the cards and also by jotting down the time required for a task's completion. He scores the first three series either right or wrong, but allows partial credits for imperfect, yet "sensible" arrangement

---

[10] Rapaport, David, Diagnostic Psychological Testing, Chicago: The Year Book Publishers, Inc., 1945, p. 215.

on the last three sets. The final score is the sum of partial credits received on the separate items.

Test 7. Picture Completion

This test differs from most tests associated with the name "picture completion", i.e., Healy Picture Completion II, in that it does not allow the subject a multiple choice from which to select the best fitting piece, to give an unfinished picture sense upon completion. Yet, it is similar to the completion test employed by Pintner in his Non-Language Scale, in which he demands that the subject draw in the missing part. Wechsler simply asks that the subject discover what is missing. But, Pintner devised his test in such a way that the actual workings of the test problems could be given and taken without the use of the English language.

In the Picture Completion test as given on the Bellevue Scale, the subject is shown a picture, like a pig without its tail, or a door minus its knob, and asked to tell the missing part. To be able to do that,

> the subject must first know what that picture represents...he must be able to appreciate that the missing part is in some way essential either to the form or function of the object or picture. In a broad way, the test measures the ability of the individual to differentiate essential from unessential details.[11]

Fifteen cards are presented to the subject, with each of the items being exposed for no longer a period of time than fifteen seconds. The score is the number of pictures for which

---

[11]Wechsler, David, The Measurement of Adult Intelligence, Baltimore: The Williams and Wilkins Co., 1944, p. 91.

correct responses are given.

Test 8. Block Design

The Block Design test is one of three measures of visual motor coordination. It includes seven patterns preceded by the examiner assembling the blocks to make two sample designs. The blocks are those used in the Kohs Block Test, each having a blue, a white, a red, a yellow, a blue/yellow, and a red/white side. Only the red, white, and red/white sides are used in this subtest, to eliminate the possible factor of color confusion.

The two sample items and the first four test items consist of four blocks each to be placed side by side to form the design; the next two items consist of nine blocks each; and the last item consists of sixteen blocks. The blocks, in each completed design, make up a square, and all the designs show some symmetry. The figure patterns in this scale differ from those employed by Kohs in his original standardization. Wechsler effected the change to avoid reproduction of items used on other scales.[12]

The items are timed. The subject receives three score units if he passes an item within the time limit and additional time credits, from one to three, if his performance is a speedy one. The designs must be absolutely correct to receive credit.

Test 9. Object Assembly

Object Assembly is the second of three performance

---

[12] Ibid., p. 92

tests measuring visual motor coordination. It consists of three simple figure formboards: a mannikin, a profile, and a hand. The mannikin, consisting of six pieces, is similar to the one devised by Pintner, only the features have been redrawn to make them more lifelike in appearance. The profile of a woman's head, and the hand, a new formboard devised by the author, consist of seven pieces each.

The test is scored by giving attention to correctness of placement and credit for time. If all the pieces are correctly placed, the subject receives a score of six. Misplaced parts deduct one score unit from the total score. Time credits are given on the profile and hand formboards if the subject has successfully arranged all the parts, or fitted them together with only one reversal, and completed the problem in less than sixty-four seconds. The credits range from one to four score units. The sum of the raw scores on the three items plus the time credits is the total score.

Test 10. Digit Symbol

A substitution test is one of the oldest of all psychological tests. The form of digit symbol test incorporated into the Wechsler-Bellevue Scale was taken from the Army Beta. It was altered, however, with respect to the matter of time allowance for administration.

The Digit Symbol subtest requires the subject to associate certain symbols with certain other symbols. He is presented with a sample line of nine simple symbols, each printed below a number from one to nine. Sixty-seven empty squares, each under one of these numbers, constitute the test

material. The subject is asked to write, in the blank below the number, the correct symbol associated with that number in the above sample line. The speed and accuracy with which he undertakes the task set before him serves as a measure of his intellectual coordination because it infers following a sample, presented visually, and motor reproduction of the symbols.

The test is a timed one, allowing the subject ninety seconds for work. His score is the total number of symbols correctly copied, regardless of neatness and precision, but recognizibility is a "must". If a letter such as "N" is not reversed to "И" as the sample requires, the subject merits only a half raw score unit.

Chapter III

PROCEDURE

The Wechsler-Bellevue Intelligence Scale was given to fifty adolescent and adult deaf, ranging in age from fourteen to fifty-eight years. The subjects were contacted through St. Andrew's Mission (Christ Church Cathedral), Grace Lutheran Church, the Bell Club, the St. Louis Chapter of the American Red Cross, St. Joseph's Institute for the Deaf, and Central Institute for the Deaf.

With the exception of the students from St. Joseph's, the subjects came to Central Institute on a designated evening for the test. After the establishment of rapport between the examiner and the subject, a short case history was recorded, including educational attainments, present occupation, onset and cause of deafness.

As noted in the case data, Table II, twenty-one of the subjects tested were males, and twenty-nine, females. Their average age as a group was 23.8 years, being 21.5 years for the males, and 25.4 for the females. The highest grade completed in a school for the deaf was tenth grade, or the equivalent of the sophomore year at a high school for the hearing. The highest grade completed in a hearing school was the senior year at college. A list of occupations, in which deaf people are presently employed as recorded in this study's sampling, includes: art, upholstering, bookkeeping, carpentry, homemaking, chemistry, aeronautical-drafting, sculpturing, nursing, jewelry repairing, and teaching lip-reading. Their choice of work indicates a wide range of ability, varying from unskilled labor,

as might be found in a factory, to highly technical professions requiring definite skill and knowledge. Fifteen subjects were congenitally deaf, while the remaining individuals lost their hearing at ages varying from three months to thirteen years. In approximately one-fourth of the cases, the cause of deafness can be attributed to meningitis. Scarlet fever, measles, and mastoid inflammations account for other subjects being aurally-handicapped.

In the time devoted to the actual testing situation all subjects seemed very cooperative and reported how much they enjoyed it. In every case, the Verbal part was administered first, followed by the Performance part, with the complete testing time lasting for a maximum of an hour and a half.

In presenting the Verbal test to deaf subjects, the examiner found it impossible from the start to follow the instructions precisely as given in the Bellevue Manual. It directs the examiner "to read questions as stated and in order given" to the subject.[13] Oral directions can be given preceding a test situation to deaf people, but not the specific items themselves. If this procedure of presenting the Information subtest had been carried out, it would have conflicted with the test's validity. Also, it would have altered the original purpose for giving the Bellevue Scale to aurally handicapped persons. Instead of tapping the subject's range of information and serving as a good measure of intelligence, the test would have become an inadequate attempt to evaluate

---

[13] Ibid., p. 172.

the individual's lip reading ability.

The examiner altered the presentation with due consideration of the deaf person's difficulty in understanding oral statements. In the subtests - Information, Comprehension, Similarities, Arithmetic, and Vocabulary - the items were typed on small white cards. The subject read the first item in the test situation to himself, gave the answer orally, which was recorded following standard procedure by the examiner, and then proceeded on with the next question.

Table II

CASE DATA OF SUBJECTS

Subject	Age	Sex	Highest grade completed in deaf schools	Highest grade completed in hearing schools	Occupation	Age of onset of deafness	Cause recorded by subjects
A	23	M	10		jewelry	birth	
B	20	F	8	(business) 4 mos.	comptometer opr.	10 mos.	accident
C	26	M	8	4 yrs. WU Art School	artist	1 year	
D	21	F	8	9	order filler	4 years	meningitis
E	24	F	8	12	clerk	7 "	
F	23	M	8	12	mach. specialist	6 "	meningitis
G	33	F	8		file clerk	2 "	
H	21	F	8	Comptometer sch.	bkpg. machine	birth	
I	57	F	8		housewife	2½ yrs.	meningitis
J	19	M	8	10	student	2 "	
K	20	F	8		factory worker	birth	
L	20	M	8	11	student	2 yrs.	
M	21	M	8	10.6	upholsterer	6 "	meningitis
N	21	F	8		Mach. opr. inspector	3 "	
O	37	F	was educated in Canada		factory	3 mos.	
P	21	F	8		factory	3 yrs.	scarlet fever
Q	27	M	8	12	carpenter	8 "	meningitis
R	19	F	8		factory	birth	
S	58	F	never attended school		housewife	7 yrs.	
T	23	F	8	12	teacher lip-read.	7 "	meningitis
U	27	M	8	BS degree graduate WashingtonU	chemist	birth	

Subject	Age	Sex	Highest grade completed in deaf schools	Highest grade completed in hearing schools	Occupation	Age of onset of deafness	Cause recorded by subjects
V	48	F	8	9	housewife	9½ yrs.	meningitis
W	18	M	8	10	student	birth	
X	30	F	9		store clk.	9 yrs.	scarlet fever
Y	15	M	7		student	birth	
Z	19	M	8	10.6	upholsterer	14 mos.	measles
AA	20	M	8	9	upholsterer	13 "	
BB	14	F	7		student	2 yrs.	meningitis
CC	17	F	8	9	student	birth	
DD	16	M	7		student	3 wks.	measles
EE	22	F	8		film inspectress		
FF	21	F	6	12	clerical wk.	2 mos.	mastoid
GG	22	M	8	12	supervisor St.Joseph's	birth	
HH	39	F	8	12	draftsman Aero-Chart Plant	2 yrs.	
II	18	F	8	4 mos. business	comptom. operator	birth	
JJ	16	F	7		student	3 yrs.	meningitis
KK	24	F	none	12	odds & ends	7 "	abscessed ear - colds
LL	26	F	none	12	secretary	13 "	nerve deafness
MM	19	M	7		student	birth	
NN	41	M	8	BA degree Univ.Minn.	sculptor	6 mos.	meningitis
OO	16	M	8	9	student	birth	
PP	21	M	8	2	shipping clerk	birth	
QQ	17	M	7		student	3½ yrs.	scarlet fever
RR	21	M	5		assembly work	birth	

Subject	Age	Sex	Highest grade completed in deaf schools	Highest grade completed in hearing schools	Occupation	Age of onset of deafness	Cause recorded by subjects
SS	20	F	7		nursing att.	15 mos.	a fall
TT	16	F	8	1	student	3 yrs.	meningitis
UU	15	F	6		student	6 "	meningitis
VV	15	F	6		student	birth	
WW	22	F	8		bin filler (shoes)	7 mos.	meningitis
XX	20	M	8	2	student	7 mos.	a fall

Chapter IV

RESULTS

The Bellevue Scales as devised by Wechsler had not, previous to this study, been applied as a Full Scale to the adolescent and adult deaf. The Performance part had been included in the test batteries given by schools for the deaf, the St. Louis Chapter of the American Red Cross, and some social service agencies to whom deaf people have applied for vocational guidance. Reason for the omission of the Verbal Scale can be found in the following quotation:

> It is now recognized that to be a fair test of the intelligence of a deaf child, the test must not require skill in language. The directions and the performance of the test must neither of them require words.[1]

A great deal of criticism has been made of performance tests as true indices of intelligence, especially when used on students in the late adolescent years. The examiner, therefore, gave the Full Scale to deaf adults, for in every case tested, language had been developed sufficiently to enable communication between the examiner and the subject, either through speech or writing.

The intelligence quotients of the deaf as obtained from measurement on the three parts of the Bellevue Scale are listed in Table III. The Verbal Scale I.Q.s range from 62 to 126, with the mean at 94.6. The spread is wide, yet this could be expected considering that the subjects tested are severely handicapped by their language insufficiency. The Per-

---

[1] Davis, Hallowell, Hearing and Deafness: A Guide for Laymen, New York: Murray Hill Books, Inc., 1947, p. 375.

formance Scale I.Q.s range from 91 to 137, with the mean centered at 117.04. The great difference in the majority of cases between Verbal and Performance Scale I.Q.s is indicative of a deaf person's lowered ability in language comprehension. The Full Scale I.Q.s reasonably strike an average between the Verbal and Performance Scale I.Q.s. The percent of vocabulary words for which correct responses were given varies from thirteen percent to eighty-one percent. The range is extremely large, with the majority of scores below fifty percent.

Table III

INDIVIDUAL SCORES ON BELLEVUE-WECHSLER TEST

Subject	Verbal Scale I.Q.	Performance Scale I.Q.	Full Scale I.Q.	Difference between Verbal and Performance I.Q.	Percent of Vocabulary words correct
A	103	110	106	7	33
B	93	135	113	42	27
C	92	125	108	33	25
D	115	124	121	9	36
E	119	114	118	- 5	70
F	122	124	125	2	61
G	98	124	111	26	51
H	99	124	111	25	29
I	104	130	117	26	57
J	91	125	109	34	31
K	74	102	86	28	19
L	91	111	101	20	25
M	95	130	112	35	30
N	74	110	90	36	23
O	68	100	83	32	24
P	84	103	92	19	15
Q	114	125	120	11	63
R	67	111	86	44	14
S	74	98	86	24	26
T	111	118	116	7	49
U	126	137	134	11	69
V	107	110	108	3	42
W	93	91	91	- 2	32
X	109	112	111	3	81
Y	94	127	111	33	20
Z	103	128	117	25	30

Subject	Verbal Scale I.Q.	Performance Scale I.Q.	Full Scale I.Q.	Difference between Verbal and Performance I.Q.	Percent of Vocabulary words correct
AA	100	118	109	18	30
BB	90	110	100	20	23
CC	101	133	119	32	27
DD	79	120	99	41	19
EE	82	119	100	37	29
FF	99	114	106	15	45
GG	97	117	107	20	27
HH	112	126	120	14	64
II	82	126	104	44	25
JJ	83	115	98	32	23
KK	111	114	113	3	69
LL	112	128	120	16	71
MM	89	129	109	40	15
NN	115	127	122	12	63
OO	100	118	110	18	32
PP	103	117	110	14	26
QQ	90	126	109	36	21
RR	62	100	79	38	13
SS	67	91	77	24	14
TT	93	112	103	19	39
UU	72	93	80	21	14
VV	78	108	91	30	20
WW	84	121	101	37	21
XX	107	122	116	15	37
Mean Verbal I.Q.					94.60
Mean Performance I.Q.					117.04
Mean Full Scale I.Q.					105.68
Mean Difference between Verbal and Performance Scale I.Q.s					22.46
Mean Percent of Vocabulary Words Correct					35

Table IV shows a comparison of the deaf and the hearing on the three scales. The statistics for the deaf are based upon fifty cases tested for this study. The statistics for the hearing are taken on the basis of approximately three thousand cases used by Wechsler to establish norms. The correlation co-

Table IV

Comparison of Deaf and Hearing on the Bellevue Scale

	Deaf		Hearing	
	r	P.E.r	r	P.E.r
Verbal I.Q. x Performance I.Q.	.59	.092	.71	.018
Verbal I.Q. x Full Scale	.91	.024	.90	.007
Performance I.Q. x Full Scale	.51	.105	.88	.008

efficient between quotients on the Verbal and the Performance Scales for the deaf was .59±.092, whereas the correlation coefficient between quotients on the Verbal and Performance Scales for the hearing was .71 ±.018, indicating that the two scales bear a more substantial relationship to each other when used with the hearing than with the deaf. Wechsler, himself, admitted that the high correlation obtained was due in part to the heterogeneity of his sampling population as regards both age and I.Q. range, for other studies show lower correlations between the Verbal and Performance I.Q.s.[2] The low correlation of the deaf could be further explained as due to the inclusion of several subjects with a severe language handicap. The closeness of the correlation coefficients between the Verbal and Full Scales of .91 ±.024 for the deaf and .90 ±.007 for the hearing shows the high degree of relationship between the Verbal part and the Full Scale for both deaf and hearing. The correlation coefficient of .88 ±.008 between the Per-

---

[2]Wechsler, David, The Measurement of Adult Intelligence, Baltimore: The Williams and Wilkins Co., 1944, p. 124.

formance and the Full Scales for the hearing is decidedly higher than that found for the deaf. Why that large difference should exist between the two groups, and the correlation for the deaf be so low, is not completely understood, unless the total weighted score representing the sum of the five Verbal tests is greater than that of the five Performance tests respectively. A simple hypothesis, proceeding from such an explanation, might conclude that the Verbal Scale I.Q. influences the Full Scale I.Q. to a much larger degree than does the Performance I.Q. Wechsler, in expressing his opinion regarding the merits of the individual tests, adds:

> At the outset of standardization it seemed to us, in the absence of any definite knowledge, that the best assumption to make about the separate tests of the scale was that they were equally important. From a qualitative point of view, subsequent statistical analysis has not, as indeed was to be expected, fully confirmed this assumption. Some of our tests are better measures of intelligence than others.[3]

Again, the language handicap of the deaf can account for the low relationship between the Performance and Full Scales. The probable errors are low, indicating statistical reliability for the limited sampling on the Bellevue Scale.

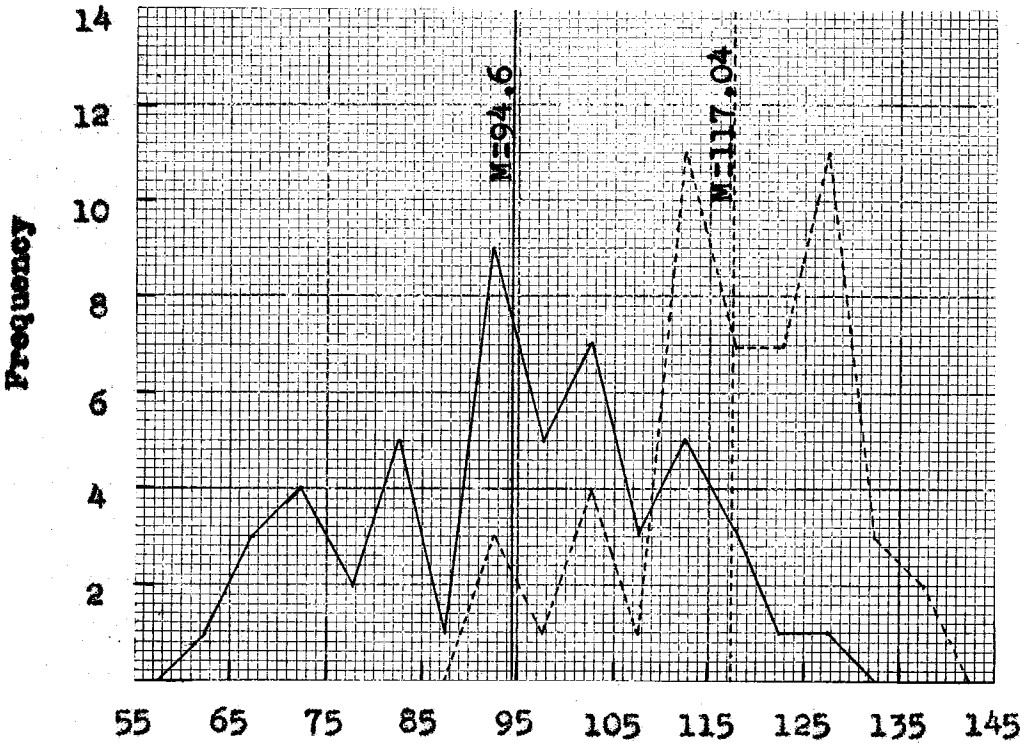
Figure II shows a comparison of the frequency distribution of the verbal and performance intelligence quotients of the deaf. The quotients are plotted in step intervals of five along the abscissae, and the frequency of occurrence of these quotients on the ordinates. Only three subjects had performance quotients lower than the mean of the Verbal Scale.

---

[3] Ibid., p. 117.

Figure II

COMPARISON OF FREQUENCY DISTRIBUTION OF PERFORMANCE AND VERBAL INTELLIGENCE QUOTIENTS



———Verbal Scale

--------Performance Scale

The average difference between the intelligence quotients on the two scales is 22.46 points. The range is wide, extending from minus five, in which case subject E's score on the Verbal Scale was five points higher than that of the Performance Scale, to plus forty-four, showing an extreme variation in ability. For example, if subject R had been given the Verbal subtests only, a final classification would have diagnosed him as mentally defective, for his quotient is computed to be only sixty-seven on the Verbal Scale. Yet, if one shares the belief that no complete measure of a person's intelligence can be made unless both performance ability and verbal ability are measured, a performance test would also be given to the subject. The above subject received an intelligence quotient of one hundred eleven on the Performance Scale, which quotient would designate the individual as having intelligence above average. That forty-four point difference is indeed large. Although it doesn't occur as the rule but rather as the exception in the comparison of the two frequency distributions, still its presence alone substantiates the fact that including both kinds of tests is a "must" when examining the deaf and the hearing.

In comparison with subject R, subject L has the same performance intelligence quotient of one hundred eleven, but his verbal intelligence quotient is higher, no doubt influenced by his educational background. He is completing his third year of high school; subject R, on the other hand, completed her education with graduation from the eighth grade and is presently employed in a factory. To what extent education,

with its constant stimulation to learning, accounts for the difference between a Verbal I.Q. of sixty-seven and one of ninety-one, is not known exactly, yet it would seem to give an examiner a more reliable measure of intelligence using a Verbal Scale.

Table V

The Influence of Education upon Intelligence Quotients

	Elementary (incomplete)	Elementary (completed)	Vocational School	High School (incomplete)	High School (completed	College completed
% of Subjects	20	22	10	22	18	4
Median Verbal I.Q.	81	84	93	98	112	120.5
Median Performance I.Q.	112.5	112	126	117	118	132
Median Full Scale I.Q.	98.5	101	111	106	118	128
Median Chronological Age	16	22	21	19-6	24	34

Table V is an attempt to show that education plays an important part in influencing a verbal intelligence quotient of the deaf. With the exception of two individuals whose education could not be accounted for either because opportunities were not present for adequate schooling or they simply didn't know what length of time they had gone to school, the subjects were classified into six scholastic groups. Twenty percent of the subjects went or are still going to elementary school. Some terminated their education early without completing the eight grades while others are striving to finish. Twenty-two percent graduated from a school for the

deaf. Only one person in that group has gone on to high school, the rest having secured employment in the business world. Ten percent, after completing grade school, took up vocational training either at a business college, a technical high school, or at Washington University Art School. Twenty-two percent of the subjects graduated from schools for the deaf, and attended high schools for the hearing, completing from a year to three years of academic work. Eighteen percent of those tested received a high school diploma, while four percent went on to college where they received their Bachelor degrees.

It is interesting to note the improving verbal intelligence quotient with amount of education as compared to the approximate constant performance intelligence quotient. The two college students are the exception in performance quotient because these quotients are extremely high for both. Perhaps, it can be explained by the fact that one is a chemist by profession, and the other, a sculptor who has won national recognition.

In presenting the Vocabulary subtest to the deaf, the examiner soon realized that the order of difficulty of the words as standardized on hearing subjects, would not be the same for the deaf. Except in a few cases where advanced education might have influenced the intelligence quotient, the approximate stopping point of an individual's knowledge of the given words could be fairly well predicted before the actual test presentation.

Final tabulation was made of the number of full points scored for each word. A subject received a full credit

Table VI

The Order of Difficulty of Vocabulary Words for Hearing and Deaf

Order of Presentation	Order of Frequency of correct response for Deaf	% of Deaf credited with correct response
1. apple	apple	100
2. donkey	donkey	100
3. join	fur	92
4. diamond	bacon	90
5. nuisance	nail	84
6. fur	gamble	82
7. cushion	join	80
8. shilling	cushion	70
9. gamble	cedar	70
10. bacon	diamond	58
11. nail	plural	56
12. cedar	shilling	42
13. tint	tint	42
14. armory	fable	38
15. fable	microscope	38
16. brim	brim	34
17. guillotine	hari-kari	32
18. plural	nitroglycerine	30
19. seclude	armory	28
20. nitroglycerine	nuisance	26
21. stanza	guillotine	24
22. microscope	stanza	24
23. vesper	belfry	20
24. belfry	pewter	20
25. recede	recede	18
26. affliction	vesper	16
27. pewter	catacomb	16
28. ballast	seclude	14
29. catacomb	affliction	12
30. spangle	spangle	12
31. espionage	espionage	12
32. imminent	imminent	10
33. mantis	mantis	8
34. hari-kari	ballast	4
35. chattel	chattel	2
36. dilatory	dilatory	0
37. amanuensis	amanuensis	0
38. proselyte	proselyte	0
39. moiety	moiety	0
40. aseptic	aseptic	0
41. flout	flout	0
42. traduce	traduce	0

of one point when his response was indicative of his knowing what the word really meant. In instances when he seemed to

know only vaguely the meaning of a word, his response was credited only as half right.

In Table III, the examiner has rearranged the test material to form a list which would be suggestive of the word-difficulty as experienced by the deaf. When the two lists are compared, it will be noted at first glance that words which are obviously very easy for hearing people to comprehend are not so understandable for the deaf. Words like join, diamond, and especially nuisance, show a higher degree of difficulty among the deaf than among the hearing. Nuisance occupies the twentieth place in order of difficulty for the deaf as compared with fifth place in order of presentation. Any acquaintance whatever which a deaf person might have with that word would come only from the teacher's orientation, and not from a "pick-up" in daily conversation. Too, it is a word which lacks concreteness. Its very abstract connotation, that of describing a state of being, halts a ready interpretation of its meaning for the deaf. Proceeding further down the list, the words fur, gamble, bacon, nail, and cedar rank as many as six places higher on the deaf list. Contrary to nuisance, four of these words have visual significance for the deaf. Being simple nouns, their understanding is easily brought about by actual contact with the objects themselves, or by pictures of such. The word gamble has undoubtedly been experienced by a majority of those tested.

Now considering the deaf list, plural, microscope, and hari-kari place high in the relative order of difficulty, meaning that they are not easier for the deaf than for the

hearing, but that the list of words presents a gradual increase of difficulty for the hearing and a "spotty" one for the deaf. The average deaf person, in the testing situation, would fail to know words, in the order presented, following tint. Having been credited with a passing definition of that word, he would "jump" to plural, then to microscope, and lastly, to hari-kari. The intervening words would be missed almost entirely, the lack of knowledge being the result of their infrequent occurrence in the daily classroom situation.

The reason that many deaf persons were able to define plural can be attributed to the fact that this word is definitely a part of every deaf child's vocabulary, beginning in primary grades. Such is not the case with a hearing boy or girl who might first learn its grammatical meaning when in the seventh or eighth grade. With the noun, hari-kari, we have a word brought to the attention of the populace with the onset of World War II. Its meaning has become clearer for hearing and deaf alike, but the test was constructed before the word came into popular use.

The most common errors made by the deaf on the test material pertain to a confusion of meaning, the result of faulty visual perception. For example, a subject would observe the word shilling, pronounce it as chilly, and supply a definition of "frightened, nervous from cold". In like manner, he would mistake the word cider for cedar, feeble for fable, T.N.T., an explosive, for tint, a dye. In the examiner's opinion, a hearing person would not make similar errors, for the word list would be presented orally. His correct pronunciation of each

word would offer no confusion so far as the subject's hearing and knowing what the actual test word is supposed to be.

In scoring the items of the vocabulary test, the examiner often met up with the difficult situation of not knowing when to give full credit. Wechsler's criteria for scoring of definitions was heeded, but his following statement couldn't always be applied wholeheartedly:

> While inelegance of expression is discarded, poverty of content is penalized to some extent.[4]

The examiner attempted to reason out the question, "Would it be fair to a deaf adult to penalize him for his scarcity of words used in language expression?". In the case of subject O, it would have been most unfair, for she defined each item with a single word. Undoubtedly, she couldn't express herself more fully. Yet, glancing at the list of words, few can be defined adequately with so sparing a quantity as one word per item. The examiner disregarded the language precision and gave her full credit for those words for which it was felt the meanings were known. This approach might have been more subjective at one time than another, especially when the examiner needed to interpret word meanings in gestures, but every individual tested had to be considered as a special clinical problem.

Table VII presents the items of the Information subtest in a shortened form. In administering the test, the full question, "What is a thermometer?" or "What does rubber come from?", was given. In the first column, the questions are

---

[4] Ibid., p. 186.

Table VII

Order of Difficulty of Information Items

Order of Presentation	Order of Relative Difficulty for Deaf	% of Deaf credited with correct response
1. President before	President before	100
2. thermometer	London	96
3. rubber	Brazil	94
4. London	thermometer	92
5. pints	rubber	92
6. weeks	pints	84
7. Italy	Japan	82
8. Japan	Washington	80
9. height	Egypt	78
10. plane	heart	78
11. Brazil	plane	76
12. Paris	Italy	72
13. heart	weeks	70
14. Hamlet	Huckleberry Finn	56
15. population	height	56
16. Washington	Paris	54
17. pole	Vatican	48
18. Egypt	Hamlet	30
19. Huckleberry Finn	pole	18
20. Vatican	population	12
21. Koran	Habeas Corpus	6
22. Faust	Koran	4
23. Habeas Corpus	Faust	0
24. ethnology	ethnology	0
25. apocrypha	apocrypha	0

listed in an order which roughly approximates their order of difficulty for Wechsler's sample of the population. In the second column, the order is rearranged to show the order of difficulty as experienced by the deaf. Questions demanding a general knowledge of geography are seemingly easy for the deaf. For example, consider the answer to the question, "Where is Brazil?". There is a difference of eight places in its position on the deaf list as compared with that on the list as presented in the test.

The results from testing the deaf on information

items confirm test data as gathered by Wechsler. He comments on some very "surprising" results with respect to the difficulty of certain of the items.

> The question, "What is the population of the United States?" turns out to be inexplicably hard. It is surprising how many native Americans do not know even the approximate number of inhabitants of their own country. We have had estimates by college graduates from 10 to 300 million. On the other hand, more people can tell what a thermometer is than how many weeks there are in a year, and more can give the name of the inventor of the aeroplane than can give the author of Hamlet.[5]

Three out of every ten deaf people were credited with a correct response to the question, "Who wrote Hamlet?". This result was not considered a poor one because of the failure of the majority to score. Rather, it showed an increasing frequency of success with higher educational levels. Those individuals whose response received credit had completed either high school or college. Only in higher education, including courses in English literature, would Shakespeare be introduced to the deaf.

Wechsler's examiners in the field have said that this test presupposes a normal or average opportunity to receive verbal information. It is a poor test for those deprived of such opportunity.[6] Would the fact that seven out of every ten deaf persons scored on half of the test items justify the latter statement so far as the deaf are concerned? To be sure, they are handicapped in language and speech acquisi-

---

[5] Ibid., p. 79

[6] Ibid., p. 80

tion, yet they strive to bridge the wide gap between formal education and incidental information by reading.

> The deaf person who reads speech and speaks well or who reads well and has good written language usually has more information and better understanding.[7]

Table VIII

Percent of Generalization of Responses on Comprehension Subtest

Test Items	Scoring Credits		
	2	1	0
1. envelope	92	2	6
2. theatre	66	14	20
3. bad company	38	36	26
4. taxes	42	34	24
5. shoes	30	64	6
6. land in city	4	52	44
7. forest	58	18	24
8. laws	10	42	48
9. marriage	12	50	38
10. deaf	18	38	44

Table VIII includes the items presented in the Comprehension subtest. The questions pertain to practical situations and considerations which the average person might experience at some time, or have to answer for himself. The percentages of those who scored on the test are listed in their respective columns: under two, where they received full credit for a correct response; under one, where their response was only half-right; and under zero, where they either failed to respond altogether, or, the answer was inadequate. This test was not an easy one for the deaf if one judges from the fact that only on three items, did the subjects show a fair

---

[7]Davis, Hallowell, Hearing and Deafness: A Guide for Laymen, New York: Murray Hill Books, Inc., 1947, p. 466.

amount of success. On questions one, two, and seven, more than half of those tested received full credit. Answers to item five, "Why are shoes made of leather?", showed a larger percentage of half credits than full credits, indicating that the context of the question itself wasn't as difficult to comprehend as the fact that it demanded more than one reason, why leather is used in the manufacture of shoes, before full credit was given. The most difficult test items were numbers six and eight. Approximately fifty percent of the subjects received no credit for their responses, while the remaining received one credit. The poor performance might be attributed to the fact that the deaf are unaccustomed to putting their ideas into words, apart from expressing themselves daily in conversation.

The variety of replies the examiner received for the last question, "Why are people who are born deaf usually unable to talk?", is far greater than one would surmise. The following are a few sample answers:

1. Because voice is lazy, don't know how to move tongue.
2. Don't have many schools for deaf to learn how to talk.
3. People are afraid to (talk), have no tongue.
4. Cannot hear.
5. Can't hear other people talking, they have not been exposed to sound atmosphere as a child.
6. If hearing and become deaf, will talk better than person born deaf.
7. A baby learns to talk by hearing his parents and repeating - deaf baby does not hear own voice.
8. They didn't learn their voice from other people.
9. Send children to state school that sign - send children too late.

Answer number one, that of an elementary student, is somewhat amusing, for it reflects the coaching which the youth has had

in speech development. No doubt, whenever her speech doesn't sound as good as it might to her instructor, she assumes that her voice is lazy. However, she missed the point of the question entirely. With response number two, the youth has rationalized the situation and placed the blame for a child's failure to talk on the lack of sufficient means for education. Response number three was a high school student's thoughts on the subject. Apparently, he didn't take in the full meaning of the question. Number four was a common reply, that couldn't be taken at face value. Many subjects who gave it seemingly didn't have any idea that an individual learns to talk by imitation of what he hears. The examiner gave full credit for answers number five and seven as their explanations were proof that the subjects had comprehended the question. Answer number six is a possible occurrence, for in all probability, a person, who has been born hearing and who suddenly becomes deaf, will speak better than a congenitally deaf person. However, this discussion and response runs afield from the test question and hence, earns no credit. The same qualifications that applied to answer number four may also be considered in the case of number eight. Lastly, number nine would seem to indicate that the subject was a firm believer in oralism; yet, his answer, too, received no credit. These multiple responses were included in this paper to show how few deaf people actually understand the reasoning behind their difficulty in learning to talk.

Table IX presents a list of paired words used for comparison in the Similarities subtest of the Verbal Scale.

Table IX

Percent of Responses to Items on Similarities Test

Test Items	Scoring Credits		
	2	1	0
1. orange-banana	58	0	42
2. coat-dress	52	10	38
3. dog-lion	68	0	32
4. wagon-bicycle	38	30	32
5. paper-radio	6	50	44
6. air-water	10	8	82
7. wood-alcohol	2	10	88
8. eye-ear	8	32	60
9. egg-seed	4	22	74
10. poem-statue	8	4	88
11. praise-punishment	0	14	86
12. fly-tree	2	4	94

With one exception, the test was free of linguistic difficulties for the deaf. The meaning of the word alcohol was not familiar to every subject tested. The method of scoring employed by Wechsler in the Bellevue Scale distinguished between superior and inferior responses. The percent of subjects who received credits of two, one, and zero, according to the discriminating scoring method, are tabulated above. If the three columns of percentages are compared with each other, it will be noted that the table is divided into approximately two groups. The place where the proportion of responses shifted from full or half credit to no credit occurs with item number six. Up to that question, a little more than fifty percent of the subjects received some credit for their replies. Here again, as with the vocabulary test, the examiner could foretell the point at which the subject would fail to score. After item number five, many of the subjects no longer found a unit characteristic belonging to both words but rather described each part of the

item in turn: i.e., instead of saying that an egg and a seed are alike in that they both are the beginning of life, they might say, "chicken can come out of egg - flower, out of seed".

The very large percent of subjects who did not score at all on items six to twelve would indicate the severe difficulty of this test for the deaf.

Table X

Comparison of Intelligence Quotients Obtained on the Bellevue Scale with Those of Other Measurements of Intelligence

Subject	Other Tests	Bellevue Verbal Scale	Bellevue Performance Scale	Bellevue Full Scale
JJ	116.6	83	115	98
MM	103.5	89	129	109
QQ	102.0	90	126	109
RR	83.3	62	100	79
XX	110.0	107	122	116
D	105.6	115	124	121
H	121.6	99	124	111
M	119.2	95	130	112
U	129.0	126	137	134
W	91.3	93	91	91
Y	126.0	94	127	111
Z	123.5	103	128	117
AA	104.7	100	118	109
BB	119.6	91	110	100
CC	132.0	101	133	119
DD	112.8	77	120	99
FF	123.0	99	114	106

Table X presents a comparison of I.Q.s of the deaf on the Advanced Performance Series and the Lectometer with the Bellevue Scale I.Q.s. The quotients computed from measurement on tests other than the Bellevue subtests were available for only seventeen subjects, as they had formerly been students at Central Institute. In forty-seven percent of the cases listed above, the scores, in the second column, closely approximate those of the Bellevue Performance Scale. This comparison is presumed to be a good one, since all three tests attempt to measure performance intelligence. In twenty-nine percent of the cases, the scores from other tests approximate the Bellevue Full Scale I.Q.s, while in twenty-four percent of the cases, they approximate the I.Q.s on the Bellevue Verbal Scale. I.Q.s from the Bellevue Scales are higher than those of other scales in thirty-one percent of the cases. The smallness of the sampling number prevents any additional comments on the relationship between the Bellevue Scales and other test scores. However, it would be interesting to note, if more cases were available, the change, if any, in I.Q.s, from the time the subjects were pupils in grade school to the time when they are adults.

Chapter V

SUMMARY AND CONCLUSIONS

1. Fifty adolescent and adult deaf were tested on the Bellevue Intelligence Scale. Comparisons were made between the Verbal and Performance Scales, between the Verbal and Full Scales, and between the Performance and Full Scales to determine the relationship between the three parts of the Bellevue Scale for the deaf, in comparison to Wechsler's standardization on the hearing. The correlation coefficient between quotients on the Verbal Scale and the Performance Scale was .59±.092 for the deaf, and .71±.018 for the hearing. Results indicate that one cannot predict from a deaf individual's score on a verbal test how he will perform on a non-verbal measure of intelligence. Contrary to that of the deaf, the correlation for the hearing is fairly high, giving evidence that a hearing person's score on a verbal test can reasonably serve as a good indicator of what his non-verbal or non-language intelligence might be.

2. Correlation coefficients between quotients on the Verbal Scale and the Full Scale were .91±.024 for the deaf and .90±.007 for the hearing. Results indicate that a high degree of relationship exists between the two scales for both the deaf and the hearing, that the Verbal Scale is a valid test of intelligence for the hearing, and that, from comparing correlation coefficients between the Verbal and Full Scales, and the Performance and Full Scales, the Full Scale is weighted more heavily with verbal items than with performance material.

3. The correlation coefficient between quotients on

the Performance Scale and the Full Scale was .51±.105 for the deaf and .88±.008 for the hearing. The Performance Scale bears a distinct relationship to the Full Scale in the measurement of hearing subjects, but not with the deaf. The Full Scale or composite intelligence quotient, in the case of the majority of deaf persons, is decreased from a high performance I.Q. because of a lower verbal I.Q.

4. The deaf subjects showed a mean intelligence quotient of 94.6 on the Verbal Scale, 117.04 on the Performance Scale, and 105.68 on the Full Scale. The average mean difference between the verbal I.Q. and the performance I.Q. was 22.46 points. The language handicap accounts for the variation which, in all probability, will ever be existent when the measurement of the intelligence of the deaf is undertaken with the aid of two widely-different testing situations such as those which a verbal and a performance scale present.

5. Classified into six successive groups depending upon the level of education attained, the subjects showed a range in intelligence quotients from a median Verbal I.Q. of 81.0 for those individuals who had not completed elementary school, to a median Verbal I.Q. of 120.5 for the two college graduates. The results give proof that a verbal intelligence quotient of the deaf can be influenced by increasing amounts of education. The performance intelligence quotients remain relatively constant, showing little if any increase with additional educational advantages.

6. The order of difficulty of vocabulary words as standardized on hearing subjects did not prove to be the same

for the deaf. The percent of words answered correctly varied from thirteen percent to eighty-one percent, with the mean located at thirty-five percent. Approximately, fifty percent of the subjects failed to score after the first eleven words had been given in the test presentation, leaving no doubt whatever that the test was a difficult one for the deaf. It definitely showed the influence of education as the vocabulary scores bore a positive relation to the verbal intelligence quotients.

7. The order of presentation of information items was rearranged to show the order of difficulty as experienced by the deaf. Approximately fifty percent of the subjects scored on seventeen out of a total twenty-five questions. In the writer's opinion, so far as adults are concerned, an information test isn't as completely inadequate a measure of intelligence for those individuals deprived of a normal opportunity to receive verbal information as popular belief assumes. Besides, results from testing the deaf confirm test data as gathered by Wechsler.

8. The Comprehension subtest was not easy for the deaf, as indicated by low scoring credits to the questions. The subjects scored well on only three out of ten items. Despite the poor performance, it was interesting to evaluate the variety of answers to one item in particular, relating to deafness. The number of subjects who clearly understood the reasoning behind their difficulty in learning to talk was surprisingly small.

9. The Similarities subtest presented a list of

twelve paired words. Barely fifty percent of the subjects received credit on the easier half of the items, and a very large percent of subjects failed completely to score on the remaining half. The severe difficulty of this test for the deaf can be attributed to their inability to make a ready transition from comprehending the concrete likenesses which two objects have for each other, to being aware of their abstract similarities.

10. In comparing the Bellevue Scale I.Q.s of the deaf with those of other measurements of intelligence, one noted, as would be expected, that the performance I.Q.s bore the closest relationship.

The results as a whole indicate normalcy in intelligence for deaf adults on both a verbal and a performance type of measurement. The adult deaf are not as retarded on a language scale as studies using verbal tests with children would indicate.

The writer feels that the Bellevue Full Scale, though weighted more with verbal than performance items, can be used in its entirety with the deaf, at age levels from fourteen to sixty years. Each individual subtest should be presented except in cases where an omission would be warranted, such as leaving out the Digit Span Test because of the subject's poor lipreading ability. In cases where deaf individuals have failed to learn speech, the Verbal section can still be presented, for they can write their answers to the questions.

The chief advantages of the scale are: (1) suitability of the tests for adults; (2) clearness in wording of the test questions; (3) interest and appeal of the subject material;

(4) ease of administration and leniency in scoring; and (5) the fact that accuracy is not sacrificed for speed on tests requiring such. The use of the tests with the deaf will give vocational guidance counselors and school administrators insight into the language comprehension of the deaf adult, as success on the Verbal Scale is dependent upon language ability.

It would be of value in further study to include the Bellevue Scale as a "regular" test in the battery of psychometric measurements used in schools for the deaf. The scale could be given when a student has completed his elementary education or when he has sufficiently matured to be considered an adult, ready to meet the challenges in the world outside his school environment. It would be interesting to correlate the results with other measures of intelligence, taken during the adult's education as a child, and to find out how constant an index of mental ability can be determined for the deaf.

The results of this study substantiate the need for adult tests to be used with the deaf as aids in vocational and educational guidance.

BIBLIOGRAPHY

Boynton, Paul L., Intelligence: Its Manifestations and Measurement, New York: D. Appleton and Co., 1933.

Brody, Leon, "Comparable Tests of Verbal and Non-Verbal Reasoning: Their Construction and Application to Developmental Problems", The Journal of Educational Psychology, 31: 180-194, 1940.

Davis, Hallowell, Hearing and Deafness: A Guide for Laymen, New York: Murray Hill Books, Inc., 1947.

Heston, Joseph C., "A Factor Analysis of Some Clinical Performance Tests", The Journal of Applied Psychology, 27: 135-149, 1943.

Rapaport, David, Diagnostic Psychological Testing, Chicago: The Year Book Publishers, Inc., 1945.

Schieffeln, Barbara, and Schwesinger, Gladys, Mental Tests and Heredity, Including A Survey of Non-Verbal Tests, New York: The Galton Publishing Co., 1930.

Stoddard, George D., The Meaning of Intelligence, New York: The Macmillan Co., 1943.

Terman, Lewis M., The Measurement of Intelligence, Boston, Mass.: Houghton Mifflin Co., 1916.

Wechsler, David, The Measurement of Adult Intelligence, Baltimore: The Williams and Wilkins Co., 1944.